Statement of
Dr. Barbara Means
Director, Center for Technology in Learning
SRI International

before the
Committee on Science, Space, and Technology
Subcommittee on Research and Science Education
United States House of Representatives

October 18, 2011

Chairman Brooks, Ranking Member Lipinski, and other members of the subcommittee, thank you for this opportunity to participate in this hearing on What Makes Successful K-12 STEM Education.

My name is Barbara Means and I direct the Center for Technology in Learning at SRI International, an independent nonprofit research organization based in Menlo Park, CA.

I was a member of the National Research Council committee chaired by Dr. Gamoran that produced the *Successful K-12 STEM Education* report.

In my testimony today, I'm going to first underline my support for what I regard as key aspects of that report and then address what I believe is the most vexing question that faces us today: Given that the federal government funds so many wonderful innovations that show promising results in the early studies, why does so little rise to the scale where it makes a real difference in schools across the country? And just to foreshadow where I'll be going, I will argue that our greatest unmet R & D need is learning how we can achieve consistently high-quality implementation of good ideas across all the variation found in American schools.

I believe that the Successful K-12 STEM Education committee's articulation of K-12 education goals not just for universal STEM literacy but for preparing broader sections of our student population for advanced-degree STEM and STEM-related occupations as well is very important. A balanced K-12 STEM education agenda will work toward all three of these goals.

And meeting these goals will require research addressing not only math and science achievement but also students' interest in STEM, their persistence in STEM courses in high school and postsecondary study, and their participation in STEM-related activities outside of school and in the job market. As Dr. Gamoran noted, we need rigorous longitudinal studies to help us understand how to develop and nurture STEM interest, persistence, and learning among student groups that now shy away from these subjects.

I am going to focus the remainder of my remarks on the steps needed to put the kinds of insights that could come from such studies into practice on a broad scale. The big challenge is scaling up what appear to be successful programs in ways that produce positive results for most or all of our students.

Conventional thinking on the part of many federal and private philanthropic programs has been that once we've identified an effective educational product or approach, we should simply roll it out to as many schools and classrooms as possible. The implicit assumption is that these schools will experience the same positive outcomes for the approach observed originally. I am going to argue that this assumption is flawed and that efforts to implement innovative K-12 STEM education approaches on a large scale need to be combined with rigorous research on those approaches in multiple contexts.

Need for Combining Scaling and Implementation Research

Educational effectiveness is a function of what gets implemented, not simply the elements of an innovation's design or a government policy.[1] And aspects of context—by which I mean factors such as grade level, school size, accountability measures, students' characteristics, and parent and community resources—have profound effects on how educational programs are interpreted and implemented.

I will illustrate this argument with the case of STEM-focused high schools. Selective STEM high schools were designed to serve our brightest students, and test scores are a major factor in gaining entrance. The bold idea behind inclusive STEM schools is to offer the same intensive focus on STEM subjects to students who are not selected by examination—to *develop* STEM expertise rather than selecting for it. It is easy to understand that instructional approaches and materials that work well with Northern Virginia's highest-scoring students who gain entrance to Thomas Jefferson High School will need to be modified in order to be effective with students who are a year or more behind national norms in math achievement when they enter an inclusive STEM high school.

Before promoting inclusive STEM high schools as a policy, we should have well-designed research demonstrating that such schools increase the likelihood that their students will be interested in, and prepared for, STEM college majors and careers. In fact, with a grant from the National Science Foundation (NSF), I am starting to examine the feasibility of conducting such a study. But this kind of research by itself is not sufficient.

If today's inclusive STEM high schools are effective, we need to figure out how we can make them widely available. For example, Texas has been particularly active in promoting inclusive STEM high schools. The Texas design for inclusive STEM schools calls for providing students with personal attention, in part by limiting school size to 100 students per grade. Although there are scores of these schools in Texas, less than 1 percent of the state's 1.4 million high school students attend them. So solid evidence that these schools are effective would lead us to the next, more difficult question. How can we obtain similar results for all of our students? The approach that works with schools of 400

---

[1] McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9,* 171-178; Spillane, J. (2004). *Standards deviation: How schools misunderstand education policy.* Cambridge, MA: Harvard University Press.

students would have to be modified for schools with 1,000 or 2,000 students, and we would not know whether it would still be effective.

The rationale for bringing a new, potentially effective educational approach to many students is obvious, but the need to support initial large-scale implementations with research is less easily understood. We tend to plan for replicating a successful education approach as if we could simply have an assembly line produce more widgets. But the components of an education approach interact with, and are shaped by, the elements of the context in which we try to implement them, as Dr. Allensworth's research illustrates. For this reason, we need to combine scaling with research on the approach as implemented under different conditions.

Over the last decade, we have invested in large-scale experimental studies to answer the question of whether certain prominent educational approaches on average produce a significant benefit. Such studies are valuable in building a knowledge base, but educators care about results for their students, not averages. And they want to know not just whether they can expect good results in their setting but *how* to implement the approach to maximize prospects for success.

Let me illustrate my point with an example that found its way into a *New York Times* article last weekend.[2] The National Evaluation of Educational Technology Interventions, of which I was a part, examined the effectiveness of 16 reading and mathematics software products implemented in grades 1, 4, 6 and high school. These particular software products were selected for this large-scale experiment because they could point to some evidence that they were effective. In the large-scale national study, however, on average, none of the products produced significantly better student achievement than was attained by students in classrooms assigned to the control condition.[3] On the other hand, for virtually every product, there were some schools in which the software-using classes out-performed the control classes, some schools where the control classes outperformed the software-using classes, and some schools where the two were equivalent. We can choose to treat such variation as random "noise," or we can focus on it as an object of study. I am among those advocating the latter stance.[4]

In the case of the national experiment on educational software, for example, we learned that features such as the students' grade level, the school's technology infrastructure, and district policies around curriculum and assessment influenced the way in which software was implemented. For example, some elementary school teachers had a set of computers in their classrooms and could have some of their students using the software while others worked with the teacher or did silent reading. Such flexibility was rare in middle and high schools where it was more common to have the whole class use the software on selected days, often in a separate computer laboratory.

---

[2] Gabriel, T., & Richtel, M. (2011). Grading the digital school: Inflating the software report card. *The New York Times*, Oct. 9, 2011.

[3] Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Means, B., et al., (2006). *Effectiveness of educational technology interventions.* Report prepared for Institute of Education Sciences, U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc.

[4] Bryk, A., Gomez, L., & Grunow, A. (2010). Getting ideas into action: Building networked improvement communities in education. Available at http://www.carnegiefoundation.org/print/7645.

The physical environment makes a difference in how an educational approach is implemented. In an extreme example, a sixth-grade class tried to use math software on laptops passed out to students in a large auditorium. The teacher could not help students because they were tightly packed in rows, so students could not get instructor assistance if they were having difficulty with the software program.

This class also provided an illustration of the inter-connected roles of teacher judgment and district policies. The math software was designed to individualize instruction, with each student working on a learning objective until he or she had mastered it. The teacher had different ideas, based upon his interpretation of school district policy. The district had instituted benchmark tests in mathematics every six weeks along with associated pacing charts indicating what should be taught in each period. In this context, the teacher felt there was no time to teach to mastery even though many of his students were English language learners who struggled with math. The infinitely patient technology tutor might have been ideal for such students, but the teacher believed that the district's policies required him to "touch upon a topic and move on."

I do not want to leave the impression that the effects of local context are always negative. Modifications of an education approach to better fit with local circumstances or the needs and interests of a particular set of students and instructors may enhance effectiveness in that setting. We found a number of examples in our studies of GLOBE, an Internet-based Earth science education program in which students took weather, vegetation, soil, and water measures for a local study site and uploaded them to a worldwide database used by both scientists and educators. Students whose teachers elaborated on the practices in the Teachers Guide by adding data analysis activities performed better than students of other GLOBE teachers on an assessment of science inquiry.[5] We found also that classes of teachers who designed extensions of the GLOBE investigations focusing on questions about their local environment were more active in the program (contributed more data to the database) than did other classrooms.[6] We brought these practices to the attention of the GLOBE program staff who were then able to build training and support for such practices into their program.

SRI spent over 10 years conducting research in support of the GLOBE program, an unusually long-lived collaboration. At the start of this joint work, the GLOBE program staff assumed that they could promote effective STEM learning activities if they simply trained teachers in how to conduct the scientific data collection protocols. They expected teachers to know how to make the data collection activities instructionally meaningful. Early on, we were able to show program staff that many teachers struggled to relate GLOBE activities to their local science curriculum. While high school teachers brought greater knowledge of science content, many of them were inexperienced in conducting hands-on activities with small groups of students. The program needed to entirely revamp its teacher training approach to address the range of needs uncovered by the research.

---

[5] Means, B., & Penuel, W. R. (2005). Scaling up technology-based educational innovations. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement*. San Francisco: Jossey-Bass.

[6] Means, B., Penuel, W. R., Crawford, V. M., Korbak, C., Lewis, A., Murphy, R. F., et al. (2001). *GLOBE Year 6 evaluation: Explaining variation in implementation.* Menlo Park, CA: SRI International.

To increase the odds that new K-12 STEM education approaches will have positive effects when implemented at a large scale, researchers should be brought in to work with educators. Researchers can contribute their expertise to implementation planning and to building in data collections that can serve as feedback for those in charge of the program. At the same time, by studying implementation in multiple contexts, researchers can advance our understanding of the necessary preconditions, critical elements, and both therapeutic and harmful adaptations of the approach.

In short, I am calling for a much closer relationship between STEM education research and K-12 STEM education practice. We need collaborative efforts aimed both at (1) scaling up approaches with prior evidence of effectiveness and (2) studying what happens in multiple settings while advising those responsible for implementing the education approaches.

Approaches to Implementation Research

In recent years the Carnegie Foundation for the Advancement of Teaching has been promoting what it calls "improvement research" incorporating design, educational engineering, and development (DEED) activity.[7] Applied to K-12 STEM education, DEED collaborations would involve scientists, researchers, and education practitioners in jointly defining a problem of practice and then developing, trying out, evaluating and revising education approaches. Repeated cycles of design, development, measurement and feedback are central to this approach.

Many of the same elements can be found in educational researchers' call for "implementation research"[8] or "design-based implementation research."[9] Defining elements of this approach are:

- a focus on important problems of educational practice as defined by practitioners and researchers,
- commitment to iterative, collaborative design,
- interest in developing a theory of program implementation through systematic inquiry, and
- concern with developing education systems' capacity for change.

Implementation research requires a kind of partnership between education research organizations and schools and districts that is rare at present, but there are several existence proofs involving mathematics or science education.[10] When the focus is STEM instructional materials, science institutions should be brought into the mix as well.

---

[7] Bryk, A., & Gomez, L. M. (2008). Ruminations on reinventing an R&D capacity for educational improvement. Available at http://www.carnegiefoundation.org/improvement-research/approach.

[8] McLaughlin, M. W. (2006). Implementation research in education: Lessons learned, lingering questions and new opportunities. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 209-228). Albany, NY: SUNY Press.

[9] Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331-337.

[10] Cobb, P. A., Henrik, E. C., & Munter, C. (2011). Conducting design research at the district level. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. Roschelle, J., Schechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J., & Gallagher, L. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal, 47*(4), 833-878. Songer, N. B., Kelcey, B., & Wenk Gotwals, A. (2009).

A key difference between the K-12 STEM education implementation research agenda I am advocating and many existing federal K-12 STEM education expenditures is the principle of striking a three-way balance between scientists, education researchers, and education practitioners. Federally funded K-12 STEM education R & D should reflect deep expertise in STEM, address problems that educators care about, and have the potential to produce generalizable insights regarding organizational change, learning, and instruction. Funded initiatives should be neither research for its own sake, nor federal underwriting of K-12 education as usual, nor feel-good programs of scientists visiting classrooms for show and tell. I am advocating long-term, sustained collaborations with the three types of partners (scientists, educators, and education researchers) having equal roles in setting the agenda.

Federal Role in K-12 STEM Education

In this country, public education is a state and local responsibility. So what role should the federal government have in K-12 STEM education? I believe that the federal government has two responsibilities in this realm. First, it can articulate our country's goals for K-12 STEM education and a vision of how to attain them. The *Successful K-12 STEM Education* report provides a starting point for articulating goals. Second, the federal government has a responsibility to support the infrastructure for improving STEM education and measuring that improvement. This infrastructure includes both concrete resources, such as assessment tools and data systems, and R & D activities, such as those I've described as implementation research. The bringing together of research and educational practice that I have described would require both intellectual and monetary investments. Individual states and districts lack the resources and the broad national vision for this undertaking.

Funding K-12 STEM Education Implementation Research

How do we fund this kind of research and implementation at scale in this time of limited resources? I am no expert in federal agency budgets, but I suspect that we could implement a significant program of K-12 STEM education implementation research using money that we are already spending that could be put to better purpose. I would look to programs that add a small K-12 education component to grants intended for STEM research activities or that add a token evaluation component to grants for STEM educational activities.

Pro forma outreach activities where a STEM professional makes a one-time visit to a classroom are unlikely to have long-term effects for education institutions, teachers or students. STEM education programs where 95% of the resources go to providing services and less than 5% to measuring whether and under what circumstances those services had positive effects are unlikely to build a robust knowledge base about how to implement effective STEM education at scale. Funding that is thinly spread across many grants and programs for "light touch" STEM education activities and perfunctory evaluations could be re-allocated toward a smaller number of significant implementation research efforts.

How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching 46*(6), 610-633.

In 2007 the Academic Competitiveness Council reported that a dozen different federal agencies were supporting 105 STEM education programs at a cost of over $3 billion ($574 million of which was for K-12 programs). Some of these are surely valuable programs, but others are likely to be too superficial to be serving our national STEM education goals. A portion of those targeting K-12 education could be consolidated or eliminated to free up funding for a significant program of K-12 STEM education implementation research.

Networks of multiple K-12 STEM education research and development collaborations, working on the same problem and sharing a common analytic framework, could accelerate the generation of knowledge about what approaches work in what contexts and with what range of implementation practices.

Policy Implications

Education approaches that are significant enough to have long-lasting consequences are necessarily complex. We need research on the resource requirements, key choices and practices in implementing K-12 STEM education approaches, and on how the approaches can be implemented to good effect in different settings.

At present, the National Science Foundation encourages proposals for implementation research under one of its field-initiated grant programs, but STEM education implementation research is not a core responsibility of any federal agency. The National Science Board Commission on 21st Century Education in STEM called attention to this gap in its 2007 national action plan (p. 14) and called for NSF to promote STEM education research on critical challenges defined by the field of educational practice.[11]

Research on STEM learning, instructional practices, and infrastructure needs to be coupled with the study of implementation and local infrastructure reform. The work needs to be designed in such a way that it both enhances the practice of participating education institutions and yields generalizable insights that build knowledge for the field. Such collaborations require new practices and new sets of skills on the part of scientists, educators, and researchers alike. Field-building activities, promoting the needed skills both in people being trained for STEM professions, education research, and education administration and in those currently engaged in these professions, will be necessary. We have seen a few isolated examples of such collaborations, but we are unlikely to see them become common without leadership and support at the federal level.

Thank you for your attention and the opportunity to submit this testimony.

[11] National Science Board. (2007). *National action plan for addressing the critical needs of the U.S. science, technology, engineering, and mathematics education system.* Arlington, VA: National Science Foundation.